Author's response to reviews

Title: Developing an algorithm to identify people with Chronic Obstructive Pulmonary Disease (COPD) using administrative data

Authors:

Margrethe Smidth (m.smidth@alm.au.dk)
Ineta Sokolowski (ineta.s@alm.au.dk)
Lone Kaersvang (lone.kaersvang@stab.rm.dk)
Peter Vedsted (p.vedsted@alm.au.dk)

Version: 2 Date: 8 March 2012

Author's response to reviews: see over
Dear Editors

We are very grateful to you for giving us the opportunity to further improve our manuscript “Developing an algorithm to identify people with Chronic Obstructive Pulmonary Disease (COPD) using administrative data” as suggested by the reviewers.

Please find below a detailed description of how we have responded to the reviewers’ suggestions and comments together with the revised manuscript.

We hope that the revised manuscript is acceptable for publication.

Yours faithfully,

Margrethe Smidth (on behalf of the authors)

PT, MSc Health Services Management, PhD fellow
The Research Unit for General Practice and
Section for General Medical Practice,
School of Public Health
Aarhus University
Bartholins Allé 2
DK-8000 Aarhus C
Denmark
Reviewer #1:

1. It is suggested that each of our three different study populations be given a specific name to easier understand which one we refer to, especially in the Discussion. We have named the three different study populations population A, population B and population C to further clarify which study population we refer to.

2. Two different prevalence rates are mentioned in the manuscript, and reviewer #1 has pointed out that we should further explain why we have chosen the lower 9% instead of the 14.3% rate that we also mention as a prevalence measure for COPD in Denmark. We have elaborated on our choice of the 9% in the Discussion as follows: “We decided to use the overall prevalence of 9% as suggested by Hansen et al. They standardized the rate to the Danish population based on their study of a stratified sample of 4,757 people out of 299,000 Danes aged 45-84 years. The 14.3% prevalence suggested by Loekke et al. was only applied to people aged 35 and above and was calculated on the basis of a study of a much smaller sample. We therefore found it relevant to use the prevalence from the study by Hansen et al in our study”.

3. Reviewer #1 correctly finds that this study is aimed at the patients already treated in the healthcare system wherefore we use data on hospitalizations, spirometry registrations and use of medication. She finds that the following sentences can be omitted: “The long-term prognosis of COPD may improve with early interventions and the earlier the patient is diagnosed the better the health gains. Organized efforts aimed at early identification for early and proactive intervention for COPD are especially important as COPD is increasingly recognized as a disease state which in principle is preventable and treatable”. The algorithm does identify the sickest patients, but also patients who might not otherwise have been identified as having COPD. We have therefore changed it as follows: “Organized efforts aimed at identification for proactive care of the patients with COPD are especially important as the health gains are better the earlier the patient is diagnosed, and the long-term prognosis of COPD may improve with early intervention (26)”.

4. Many patients with asthma may be included in the younger groups, which may explain the low identification rate as reviewer #2 suggests. To further explain this and its implications we have added the following sentence in the Discussion: “If we want to further develop the algorithm to only identify patients with COPD, we could include more variables, e.g. the International Classification of Primary Care (ICPC-2) codes, the use of which will become compulsory in general
practice in Denmark before 2013. We could decide to use the algorithm in its present form for people aged 55 or older as it has the best properties in this group”.

5. As recommended by reviewer #1 we have changed line 3 in the GP verification paragraph, the Material and methods section, to: “The GPs were asked to refer to the GOLD guidelines where the diagnosis of COPD is confirmed by means of post-bronchodilator spirometry which shows a forced expiratory volume in one second (FEV1)/forced vital capacity (FVC)<70%”.

6. We have stated that: “A very high specificity of the algorithm is one of its strengths. This ensures that less healthy people will be suspected of having COPD unnecessarily.” Reviewer #1 questions if this is a legitimate statement. We have based the algorithm on data on hospitalization, use of medication and spirometry registration and have no knowledge of the identified patients’ diagnoses. The algorithm was developed for the purpose of identifying patients to be properly classified for proactive care. We have accordingly changed it to: “A very high specificity of the algorithm is one of its strengths. It ensures that people who have been hospitalized with a lung-related diagnosis, been prescribed lung-related medication or had a spirometry test for other diagnoses will not unnecessarily be suspected of having COPD”.

7. “… the GOLD Standard for COPD” in the Discussion section, paragraph 6 has been replaced with “the GOLD guidelines for COPD” as correctly suggested by the reviewer.

8. Reviewer #1 asks us to elaborate on: “This could both mean that less and that more patients could have been identified” in the Discussion, paragraph 6. We have added the following just after the next sentence: “We decided to use the overall prevalence of 9% as suggested by Hansen et al. They standardized the rate to the Danish population based on their study of a stratified sample of 4,757 people out of 299,000 Danes aged 45-84 years. The 14.3% prevalence suggested by Loekke et al. only applied to people aged 35 and above and was calculated on the basis of a study of a much smaller sample. We therefore found it relevant to use the prevalence from the study by Hansen et al in our study”. The next sentence has been changed to: “Probably more, considering the fact that it has been suggested that the overall prevalence of people with COPD is 14.2%”.

9. A missing punctuation mark has been included between “disease” and “Thus” in the Discussion, paragraph 6.
10. Reviewer #1 has kindly suggested amending the List of abbreviations by adding “Pulmonary” to “Chronic Obstructive Disease” and by correcting the spelling mistake in “SD” to “Standard Deviation”, which we have done.

11. We have added a column to the left in Table 3 with the 10-year age groups as reviewer #1 has recommended.

Reviewer #2:

1. As reviewer #2 correctly states, this study has three steps. It is, however, different from a serial testing as each step has its own study population. To differentiate between them reviewer #1 suggested naming them, which we have done. We hope this clarifies the different calculations. Furthermore, we have added a table - now Table 3 – showing the sensitivity, specificity, PPV and NPV for the five different GP surgeries when adding the “not certain” answers to the certain verification of the COPD diagnosis of the identified patients.

2. We have deleted the sentence “... the algorithm should mostly be regarded as a tool for identifying chronic lung disease” as suggested by reviewer #2.

3. Reviewer #2 states that the aim of our study was to identify as many patients as possible, and thus the sensitivity is the most important measure. The overall sensitivity for population C was 29.7 and for population B 44.8. We hope to have stressed this further by adding a table – now table 3.